ANALISIS ALGORITMA *NAIVE BAYES CLASSIFIER* DENGAN SELEKSI ATRIBUT BERBASIS *PARTICLE SWARM OPTIMIZATION* UNTUK KLASIFIKASI PELANGGAN *CROSS SELLING* PRODUK INDIHOME PT. TELKOM JAKARTA

RESUME TESIS

Oleh

INDAH PURNAMASARI

NIM : 14001240

PROGRAM PASCA SARJANA MAGISTER ILMU KOMPUTER

SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN KOMPUTER

NUSAMANDIRI

JAKARTA

2016

HALAMAN PENGESAHAN

HALAMAN PENGESAHAN

Tesis ini diajukan oleh :

Nama	: Indah Purnamasari
NIM	: 14001240
Program Studi	: Magsiter Ilmu Komputer
Jenjang	: Strata Dua (S2)
Konsentrasi	: *Management Information System* (MIS)
Judul Tesis	: "Analisis Algoritma *Naive Bayes Classifier* dengan seleksi atribut berbasis *Particle Swarm Optimization* untuk Klasifikasi Pelanggan *Cross selling* Produk Indihome PT. Telkom Jakarta"

Telah berhasil dipertahankan dihadapan Dewan Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Magister Ilmu Komputer (M.Kom) pada Program Pascasarjana Magister Ilmu Komputer Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri (STMIK Nusa Mandiri).

Jakarta, 20 Febuari 2016
Pascasarjana Magister Ilmu Komputer
STMIK Nusa Mandiri
Direktur

Prof. Dr. Ir. R. Eko Indrajit, M.Sc, MBA

D E W A N P E N G U J I

Penguji I	: Dr. Said Mirza Pahlevi, M.Eng.	.....
Penguji II	: Dr. Windu Gata, M.Kom.	.....
Penguji III / Pembimbing	: Dr. Sfenrianto, M. Kom.	.....

LEMBAR BIMBINGAN

STMIK - STIBA NUSA MANDIRI	**Kartu Bimbingan Tesis** **STMIK Nusa Mandiri Jakarta**

Nama : Indah Purnamasari

NIM : 14001240

Judul Tesis : ANALISIS ALGORITMA *NAIVE BAYES CLASSIFIER* DENGAN SELEKSI ATRIBUT BERBASIS *PARTICLE SWARM OPTIMIZATION* UNTUK KLASIFIKASI PELANGGAN *CROSS SELLING* PRODUK INDIHOME PT. TELKOM JAKARTA

No.	Tanggal Bimbingan	Pokok Bahasan	Paraf Dosen Pembimbing
1.	7 November 2015	Bimbingan Perdana Pengajuan Judul Tesis	
2.	21 November 2015	Bimbingan Data Tesis	
3.	5 Desember 2015	Bimbingan BAB I	
4.	19 Desember 2015	ACC BAB I dan Revisi BAB II	
5.	9 Januari 2016	ACC BAB II dan Revisi BAB III	
6.	23 Januari 2016	ACC BAB III dan Revisi BAB IV & V	
7.	6 Febuari 2016	ACC BAB I – V	

Dimulai pada tanggal : 7 November 2015

Diakhiri pada tanggal : 6 Febuari 2016

Jumlah Pertemuan Bimbingan : 7 Pertemuan

Disetujui oleh,

Dosen Pembimbing

(Dr. Sfenrianto, M.Kom.)

ABSTRAKSI

Salah satu strategi pemasaran PT. TELKOM yaitu dengan *cross selling.* Selain untuk memperoleh keuntungan dari produk tambahan yang dijual, juga untuk meningkatkan ketergantungan pelanggan pada vendor sehingga mengurangi *churn*. Pemasaran *cross selling* dapat dilakukan dengan menawarkan produk baru. Tetapi terlalu banyak penawaran terhadap pelanggan yang tidak tepat, menjadikan pemasaran tidak efisien dan efektif. Untuk itu perlu dilakukan klasifikasi pelanggan telepon dan speedy eksisting PT. TELKOM Jakarta untuk *cross selling* produk Indihome. Penelitian ini menggunakan *data mining* klasifikasi dengan algoritma *Naive Bayes Classifier* dan seleksi atribut *Particle Swarm Optimization* untuk meningkatkan akurasi klasifikasi dari 85,08 % menjadi 89,31%.

Kata kunci : *Naive Bayes Classification (NBC), Particle Swarm Optimization (PSO),* Indihome, *Data Mining*

ABSTRACT

One of PT. TELKOM marketing strategy is by cross selling. In addition to the benefit of additional products sold, as well as to enhance the customer's dependence on vendors, thereby reducing churn. Marketing cross selling can be done by offering new products. But too many deals to customers who are not appropriate, make marketing inefficient and ineffective. It is necessary for the classification of existing telephone customers and speedy PT. TELKOM Jakarta *for cross selling Indihome. This study uses data mining classification with Naive Bayes classifier algorithm and selection of attributes Particle Swarm Optimization to improve the classification accuracy of 85,08% to 89,31%.*

Keyword : Naive Bayes Classification (NBC), *Particle Swarm Optimization* (PSO), *Indihome, Data Mining*

DAFTAR ISI

Halaman
HALAMAN SAMPUL i
HALAMAN PENGESAHAN ixi
ABSTRAK iii
DAFTAR ISI xii
BAB 1 PENDAHULUAN 1
BAB 2 LANDASAN TEORI 2
2.1 Tinjauan Pustaka 2
2.1.1 Data Mining 2
2.1.2 Naive Bayes Classifier *(NBC)* 2
2.1.3 Particle Swarm Optimization (PSO) 2
2.1.4 Evaluasi dan Validasi 3
2.1.5 Indihome 4
2.2 Tinjauan Studi 5
2.3 Tinjauan Objek Penelitian 7
BAB 3 METODOLOGI PENELITIAN 8
3.1 Desain Penelitian 8
BAB 4 HASIL ANALISA DAN PEMBAHASAN 9
4.1 Pengumpulan Data 9
4.2 Preprocessing Data 10
4.3 Pendekatan *Data Mining* 12
4.3.1 Algoritma *Naive Bayes Classifier* 12
4.3.2 Algoritma *Particle Swarm Optimization - Naive Bayes Classifier* ... 12
4.4 Eksperimen 12
4.4.1 Algoritma *Naive Bayes Classifier* 14
4.4.2 Algoritma *Particle Swarm Optimization-Naive Bayes Classifier* 14
4.5 Pengukuran 16
4.5.1 Algoritma *Naive Bayes Classifier* 16
4.5.2 Algoritma *Particle Swarm Optimization - Naive Bayes Classifier* 16
4.6 Hasil Analisa 17
4.6.1 Algoritma *Naive Bayes Classifier* 17
4.6.2 Algoritma *Particle Swarm Optimization - Naive Bayes Classifier* 17
4.7 Pembahasan 18
4.7.1 Peningkatan akurasi 18
4.7.2 Uji *T-test* 20
4.8 Penerapan 20
BAB 5 PENUTUP 22
5.1 Kesimpulan 22
5.2 Saran 22
DAFTAR REFERENSI ..... **Error! Bookmark not defined.**

BAB 1 PENDAHULUAN

PT. Telkom sebagai perusahaan yang bergerak di bidang telekomunikasi dan informasi di Indonesia memperkenalkan layanan *Internet Protocol Television* (IPTV) yaitu metode penyaluran gambar dan suara televisi melalui IP pada tahun 2011 dengan nama produk Groovia TV.

Strategi pemasaran terhadap pelanggan yang telah ada (eksisting) disebut dengan *cross selling. Cross-selling* adalah strategi menjual produk baru untuk pelanggan yang telah melakukan pembelian lainnya sebelumnya. Selain untuk memperoleh keuntungan dari produk tambahan yang dijual, juga meningkatkan ketergantungan pelanggan pada vendor sehingga mengurangi churn (Jaroszewicz, 2008). Pemasaran untuk pelanggan baru akan lebih mahal daripada mempertahankan pelanggan setia yang sudah ada (Sharma & Panigrahi, 2011). Oleh karena itu, perusahaan menjual produk baru terhadap pelanggan yang telah ada untuk meningkatkan keuntungan perusahaan.

Pemasaran *cross selling* dapat dilakukan dengan menawarkan produk baru melalui SMS, panggilan telepon ke pelanggan dan surat ke pelanggan bersamaan dengan tagihan bulanan. Akan tetapi terlalu banyak penawaran terhadap pelanggan yang tidak tepat, hanya akan membuat kejenuhan di sisi pelanggan dan akan berdampak negatif yaitu pelanggan menolak produk yang ditawarkan (Jaroszewicz, 2008). Untuk itu agar pemasaran produk Indihome dapat lebih efisien dan efektif maka dilakukan klasifikasi kepada pelanggan telepon dan pelanggan speedy eksisting PT. Telkom Jakarta yang berpotensi untuk meningkatkkan layanannya menjadi pelanggan Indihome.

Banyak penelitian yang telah dilakukan dalam klasifikasi dengan algoritma *Naive Bayes Classifier* (NBC) diantaranya yaitu penelitian yang menyatakan bahwa NBC membutuhkan waktu lebih sedikit untuk membangun model (Karim & Rahman, 2013).

Akan tetapi NBC memiliki kelemahan utama pada asumsi fitur *independent* sehingga dengan menggunakan seleksi atribut dapat meningkatkan kinerja *classifier* (Taha et al., 2013).

BAB 2 LANDASAN TEORI

2.1 Tinjauan Pustaka

2.1.1 Data Mining

Data mining adalah proses menemukan korelasi bermakna baru, pola dan trend oleh sejumlah data besar yang disimpan dalam repositori menggunakan teknologi pola serta teknik statistik dan matematika (Larose, 2005, p. 2). *Data mining* adalah proses analitik dirancang untuk mengeksplorasi sejumlah besar data mencari ilmu tersembunyi yang konsisten dan berharga, langkah pertama terdiri dalam eksplorasi dan data persiapan awal (Gorunescu, 2011, p. 57). *Data mining* didefinisikan sebagai proses menemukan pola dalam data(Witten et al., 2011, p. 5).

2.1.2 Naive Bayes Classifier *(NBC)*

Naive Bayes Classifier adalah salah satu algoritma klasifikasi data mining untuk menemukan model atau fungsi yang menjelaskan atau membedakan kelas data. Teorema Bayes dikemukakan oleh Reverand Thomas Bayes pada abad ke 18 (Bramer, 2007, p. 24-28). Pada algoritma *Naïve Bayes* (NB) didasari oleh Persamaan Teorema Bayes, yang membedakan adalah asumsi keindependenan antara atribut. Persamaan Bayes adalah (Gorunescu, 2011, p. 186-190):

$$\boldsymbol{P}(\boldsymbol{H}|\boldsymbol{X}) = \frac{\boldsymbol{P}(\boldsymbol{X}|\boldsymbol{H}).\ \boldsymbol{P}(\boldsymbol{H})}{\boldsymbol{P}(\boldsymbol{X})} \qquad (2.1)$$

Untuk menjelaskan teorema *Naive Bayes*, perlu diketahui bahwa proses klasifikasi memerlukan sejumlah petunjuk untuk menentukan kelas apa yang cocok bagi sampel yang dianalisis tersebut. Karena itu, teorema Bayes di atas disesuaikan sebagai berikut (Gorunescu, 2011, p. 186-190):

$$\boldsymbol{P}(\boldsymbol{H}|\boldsymbol{X}) = \boldsymbol{P}(\boldsymbol{X}|\boldsymbol{H}).\ \boldsymbol{P}(\mathbf{H}) \qquad (2.2)$$

2.1.3 Particle Swarm Optimization (PSO)

Metode PSO merupakan teknik optimasi yang dikembangkan oleh Kennedy dan Eberhart pada tahun 1995. *Particle Swarm Optimization* merupakan algoritma

yang sangat sederhana yang efektif untuk mengoptimalkan berbagai fungsi. (Kennedy & Eberhart, 1995). Partikel memperbaharui kecepatan (*velocity*) dengan persamaan:

$$\boldsymbol{v_{id}} = \boldsymbol{\omega} * \boldsymbol{v_{id}} + \boldsymbol{\eta_X} * \boldsymbol{R}(\quad) * (\boldsymbol{p_{id}} - \boldsymbol{x_{id}}) + \boldsymbol{\eta_2} * \boldsymbol{R}(\quad) * (\boldsymbol{p_{gd}} - \boldsymbol{x_{id}}) \quad (2.3)$$

$$\boldsymbol{x_{id}} = \boldsymbol{x_{id}} + \boldsymbol{v_{id}} \quad (2.4)$$

2.1.4 Evaluasi dan Validasi

Evaluasi dan validasi untuk menguji model pada penelitian ini menggunakan *cross validation, confusion matriks, ROC curve, T-test.*

1. *Cross Validation*

   Validasi merupakan proses evaluasi untuk mengukur keakuratan prediksi suatu model. Tahap validasi mewakili langkah paling penting dalam proses membangun sebuah model (Bramer, 2007, p. 82).

2. *Confusion Matriks*

   Confusion matrix memberikan keputusan yang diperoleh dalam traning dan testing. *Confusion matrix* berisi informasi aktual (actual) dan prediksi (predicted) pada sistem klasifikasi. Rumus yang berkaitan dengan perhitungan akurasi adalah sebagai berikut (Gorunescu, 2011, p. 320) :

Tabel 2.1 *Confusion Matrix*

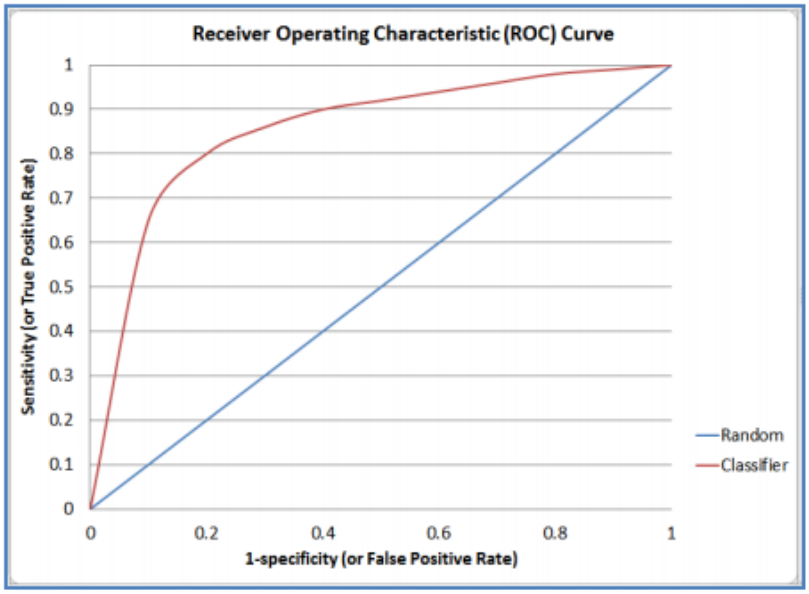
CLASSIFICATION	*PREDICTED CLASS*	
	Class=Yes	*Class=No*
Class=Yes	*A (True Positive - tp)*	*B (False Negative - fn)*
Class=No	*C (False Positive - fp)*	*D (True Negative - tn)*

Berikut adalah persamaan model *confusion matrix* untuk menghitung akurasi. Nilai akurasi adalah proporsi jumlah prediksi yang benar. Dapat dihitung dengan menggunakan persamaan (Kohavi and John, 1997).

$$\text{Akurasi} = \frac{tp + tn}{tp + tn + fp + fn}. \quad (2.5)$$

3. *ROC Curve*

Curve ROC (Receiver Operating Characteristic) adalah cara lain untuk mengevaluasi akurasi dari klasifikasi secara visual. Sebuah grafik ROC adalah plot dua dimensi dengan proporsi positif salah (fp) pada sumbu X dan proporsi positif benar (tp) pada sumbu Y.



Gambar 2.1 ROC *Curve*

4. *T-test*

Analisis *Paired-Samples T Test* merupakan prosedur yang digunakan untuk membandingkan rata-rata dua variabel dalam satu group. Artinya pula analisis ini berguna untuk melakukan pengujian terhadap dua sampel yang berhubungan atau dua sampel berpasangan.(Elliott & Woodward, 2007)

2.1.5 Indihome

Indihome Internet Fiber merupakan layanan *Triple Play* dari PT. TELKOM yaitu suatu produk terbaru PT. TELKOM yang terdiri dari 3 layanan sekaligus antara lain *High Speed Internet on Fiber* (Internet cepat), *Interactive* TV (Usee TV) dan Telepon Rumah dengan menggunakan 100% fiber optic artinya fiber optic digelar sampai ke rumah pelanggan (*Fiber To The Home /* FTTH). Indihome juga memiliki beberapa atribut tambahan seperti wifi.id *Seamless*, Indihome Telkomsel

Mania, Indihome *Global Call*, Indihome *view*, MelOn dan *Trend Micro Security System*.

Gambar 2.2 Layanan *Triple Play* Indihome

2.2 Tinjauan Studi

Penerapan data mining yang digunakan untuk penjualan produk lain (*cross selling*) terhadap pelanggan eksisting telah dilakukan penelitian sebelumnya, diantaranya yaitu:

Tabel 2.2 Penelitian yang Relevan

No.	Nama Peneliti	Judul Penelitian	Masalah Penelitian	Metode	Kesimpulan
1.	Hyunchul Ahn, Jae Joon Ahn, Kyong Joo Ohb, Dong Ha Kim	*Facilitating cross-selling in a mobile telecom market to develop customer classification model based on hybrid data mining techniques*	Mendapatkan algoritma DM yang dapat menyelesaikan klasifikasi pelanggan untuk	logistik regresi, jaringan saraf tiruan, dan pohon keputusan dengan Genetic Algorithm (GAOW)	GAOW menghasilkan akurasi training data set 66,60% dan akurasi hold out dataset 66,02%

			layanan baru		
No.	**Nama Peneliti**	**Judul Penelitian**	**Masalah Penelitian**	**Metode**	**Kesimpulan**
2.	Szymon Jaroszewicz	*Cross-selling models for telecommunication Services*	Mendapatkan algoritma DM yang dapat menyelesaikan klasifikasi pelanggan untuk layanan baru	Algoritma NB, DT, J4.8, SVM	Akurasi terbaik menggunakan SVM 28,15%
3.	Navid Khozein Ghanad,Saheb Ahmadi	*Combination of PSO Algorithm and Naive Bayesian Classification for Parkinson Disease Diagnosis*	Mencari metode klasifasi	PSO dan NB	Kombinasi metode PSO dan NB memperoleh akurasi data pelatihan tertinggi 97.95%
4.	Masud Karim, Rashedur M. Rahman	*Decision Tree and Naïve Bayes Algorithm for Classification and Generation of Actionable Knowledge for*	Mendapatkan algoritma DM yang dapat menyelesaikan	Algoritma NB dan C4.5	Akurasi C4.5 yaitu 0,94 lebih baik dari NB namun NB lebih cepat yaitu 0.58 untuk membangun model

		Direct Marketing	secara efisien		
No.	**Nama Peneliti**	**Judul Penelitian**	**Masalah Penelitian**	**Metode**	**Kesimpulan**
5.	Sergio Moro, Raul M. S. Laureano dan Paulo Cortez	*Using Data Mining For Bank Direct Marketing :An Application Of The CRISP-DM Methodology*	Meningkatkan efisiensi pemasaran menggunakan DM terbaik	CRISP-DM dengan algoritma NB, DT, SVM	NB dapat digunakan pada database besar dengan 59 atribut. SVM mencapai kinerja tertinggi AUC 0.938 dan ALIFT 0.887
6.	Hany A. Elsalamony	*Bank Direct Marketing Analysis of Data Mining Techniques*	Mencari algoritma klasifikasi untuk prediksi pemasaran	MPLNN, TAN, LR, C5.0	Akurasi terbaik dihasilkan oleh MPLNN 90.49%.

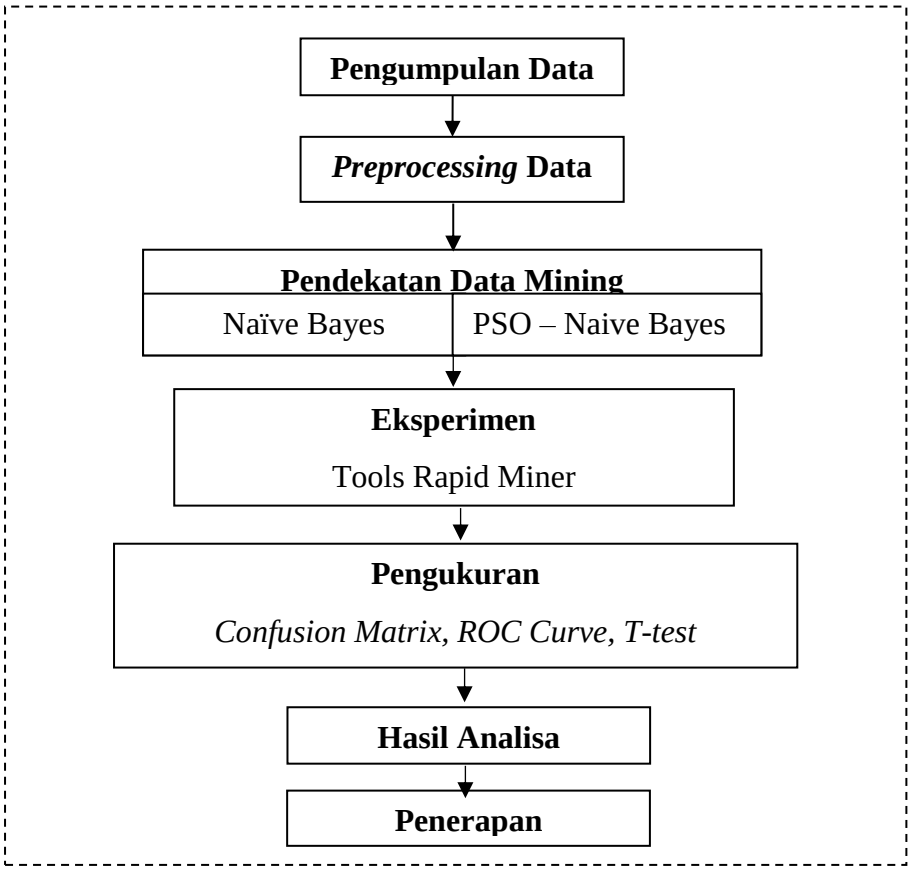
2.3 Tinjauan Objek Penelitian

Objek penelitian yang akan digunakan sebagai data dalam penelitian ini adalah data primer pelanggan telepon dan speedy *eksisting* dari divisi *Bussiness Planning and Performance* PT. Telkom Jakarta.

BAB 3 METODOLOGI PENELITIAN

3.1 Desain Penelitian

Tahapan dalam penelitian ini yaitu sebagai berikut :



Gambar 3.1 Tahapan Penelitian

BAB 4 HASIL ANALISA DAN PEMBAHASAN

4.1 Pengumpulan Data

Data yang diperoleh merupakan data primer dari divisi *Bussiness Planning and Performance* PT. Telkom Jakarta yang terdiri atas 878 data pelanggan telepon dan speedy eksisting dengan 14 atribut termasuk *class label atribute* (atribut output) yaitu atribut Paket. Data-data ini merupakan data fakultatif yang telah diketahui statusnya yaitu 3P (pelanggan Indihome) atau *decline* (menolak penawaran produk Indihome). Dari 878 data pelanggan telepon dan speedy eksisting terdapat 587 pelanggan Indihome (3P) dan 291 pelanggan telepon dan speedy yang menolak penawaran Indihome (*Decline*). 14 atribut yang terdapat pada data pelanggan telepon dan speedy eksisting tersebut adalah :

1. NCLI : Nomor identitas pelanggan
2. Nama : Nama pelanggan
3. Alamat : Alamat pelanggan
4. Regional : Regional 2 PT. TELKOM Jakarta meliputi Jabotabes (Jakarta, Bogor, Tangerang, Bekasi, Serang)
5. Witel : Wilayah telepon (Jakarta Utara, Jakarta Barat, Jakarta Selatan, 7 Jakarta Timur, Jakarta Pusat, Banten Barat (Serang), Banten Timur (Tangerang), Jabar Utara (Bekasi), Jabar Barat (Bogor))
6. POTS : *Plain Old Telephone Service* merupakan kode wilayah dan nomor telepon pelanggan (21xxxxxxxx, 251xxxxxxxx, 254xxxxxxxxx)
7. Speedy : Nomor identitas pelanggan speedy
8. Kwadran : Ada pemakaian maka ada tagihan (4)
9. Paket Eksisting : Jenis paket layanan yang saat ini digunakan (INETL15H, INETL50H, INETU1M, INETU2M,

	INETU384K, INETU384R, INETU3M, INETU512K, INETU512R, INET_1MB_H, INET_2MB_H, INET1000S, INETG10M1, INETG2M1, INETR10M1, INETR10M3, INETR1M1, INETR2M1, INETR2M3, INETR3M1, INETR3M2, INETR3M3, INETR50M1, INETR512K1, INETR512K2, INETR512K3, INETR5M1, INETR5M3)
10. R2BB	: *Ready To Broadband* (GIPTV, GSP5, GTLP, IPTV, SP2, SP3, SP4, SP5, SP7)
11. Tipe Jaringan	: Tipe Jaringan yang digunakan saat ini (*Copper*, FTTH, MSAN)
12. Zona	: Zona waktu pemasaran (*attack_promo, super_winning, winning)*
13. Migrasi	: Jenis pelanggan saat ini (2p = pelanggan telepon dan speedy)
14. Paket	: Apakah pelanggan akan meningkatkan layanan menjadi pelanggan indihome (decline, 3P)

4.2 Preprocessing Data

Data ini memiliki 14 atribut termasuk *class label atribute* (atribut output) yaitu atribut Paket. Akan tetapi 5 atribut NCLI, Nama, Alamat, POTS, Speedy mempunyai nilai yang tidak sama (unik) untuk setiap pelanggan sehingga pada penelitian ini hanya menggunakan 9 atribut

Selanjutnya data pelanggan telepon dan speedy eksisting PT. TELKOM Jakarta yang diperoleh selanjutnya ditinjau kelengkapan datanya. Dari data yang diperoleh diketahui bahwa terdapat *missing value* pada atribut kwadran dan paket eksisting pada dataset pelanggan telepon dan speedy eksisting ke 830, 838 dan 842.

Tabel 4.1 Missing Value data

NO	PAKET	REGIONAL	WITEL	KWADRAN	PAKET EXISTING	R2BB	TYPE JARINGAN	ZONA	MIGRASI
830	3P	2	JAKARTA UTARA	?	GIPTV	SP3	FTTH	SUPER_WINNING	2P
838	3P	2	JAKARTA BARAT	?	INETR2M3	SP4	COPPER	SUPER_WINNING	2P
842	3P	2	JAKARTA BARAT	?	?	SP4	COPPER	SUPER_WINNING	2P

Oleh karena itu dilakukan *replace missing value* pada data dengan *missing value* dengan nilai rata-rata. Hasil Replace Missing Value ditunjukkan pada tabel berikut ini :

Tabel 4.2 Hasil *Replace Missing Value*

NO	PAKET	REGIONAL	WITEL	KWADRAN	PAKET EXISTING	R2BB	TYPE JARINGAN	ZONA	MIGRASI
830	3P	2	JAKARTA UTARA	4	GIPTV	SP3	FTTH	SUPER_WINNING	2P
838	3P	2	JAKARTA BARAT	4	INETR2M3	SP4	COPPER	SUPER_WINNING	2P
842	3P	2	JAKARTA BARAT	4	INETR1M1	SP4	COPPER	SUPER_WINNING	2P

Tabel diatas menunjukkan hasil *Replace Missing Value* dimana data ke 830, 838 dan 842 yang pada awalnya terdapat *missing value* telah terisi dengan nilai rata-rata yang ada dimana nilai atribut untuk :

- Data 830
  kwadran : 4
- Data 838
  kwadran : 4
- Data 842
  kwadran : 4
  paket eksisting : INETR1M1

4.3 Pendekatan *Data Mining*

4.3.1 Algoritma *Naive Bayes Classifier*

Data-data yang digunakan telah diketahui statusnya yaitu 3P (pelanggan Indihome) atau *decline* (menolak penawaran produk Indihome). Dari 878 data pelanggan telepon dan speedy eksisting terdapat 587 pelanggan Indihome (3P) dan 291 pelanggan telepon dan speedy yang menolak penawaran Indihome (*Decline*) sehingga dengan menggunakan persamaan 2.1, maka :

P (3P) = 587 : 878 = 0,66856492

P (Decline) = 291 : 878 = 0,33143508

4.3.2 Algoritma *Particle Swarm Optimization - Naive Bayes Classifier*

Dari hasil seleksi atribut diperoleh bahwa atribut Witel dan Tipe Jaringan bernilai lebih dari 0. Hal ini menunjukkan bahwa atribut-atribut yang bernilai lebih dari 0 memiliki keterkaitan terhadap klasifikasi suatu data sehingga digunakan sebagai atribut prediksi dalam klasifikasi pelanggan telepon dan speedy *cross seling* produk Indihome. Dengan menggunakan persamaan 2.1, maka :

P (3P) = 587 : 878 = 0,66856492

P (Decline) = 291 : 878 = 0,33143508

4.4 Eksperimen

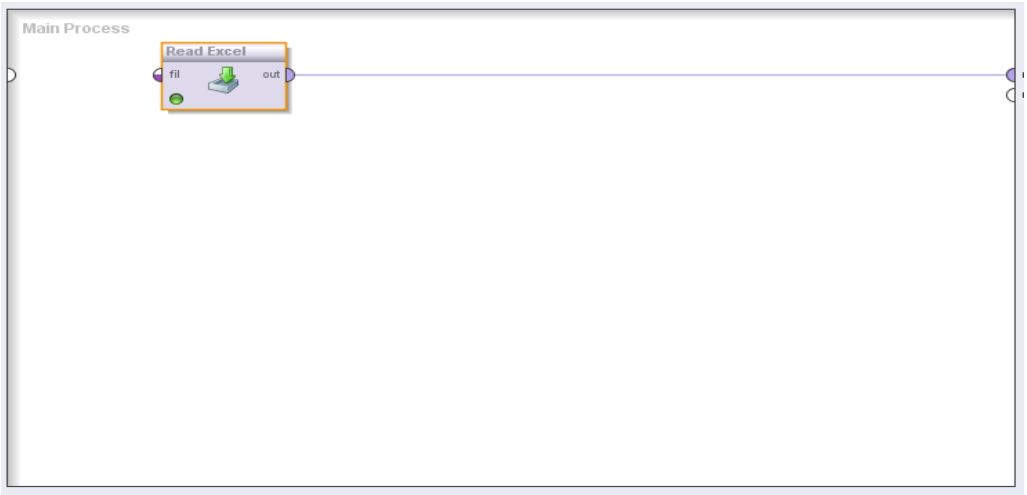
Eksperimen yang dilakukan mengunakan alat bantu untuk pemrosesan data dengan spesifikasi *hardware* dan *software* sebagai berikut :

Tabel 4.3 Spesifikasi *hardware* dan *software*

Hardware	***Software***
CPU : Intel Core i3 (1,8GHz)	Sistem Operasi : Windows 7
Memory : 2 GB	Data Mining : Rapidminer versi 5.3
Harddisk : 500 GB	

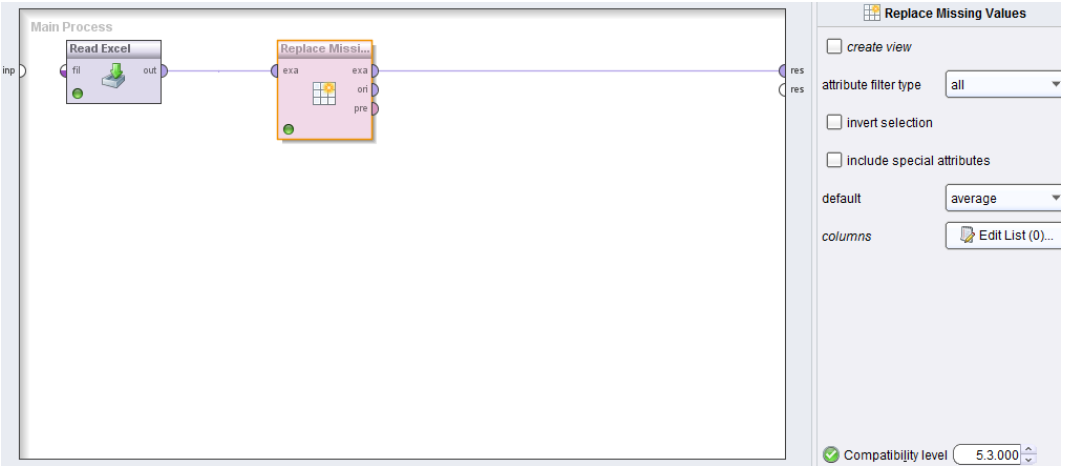
Penggunaan software Rapidminer 5.3

Untuk melihat kelengkapan data agar mengetahui *missing value* maka dilakukan proses sebagai berikut :



Gambar 4.1 Cara melihat *missing value*

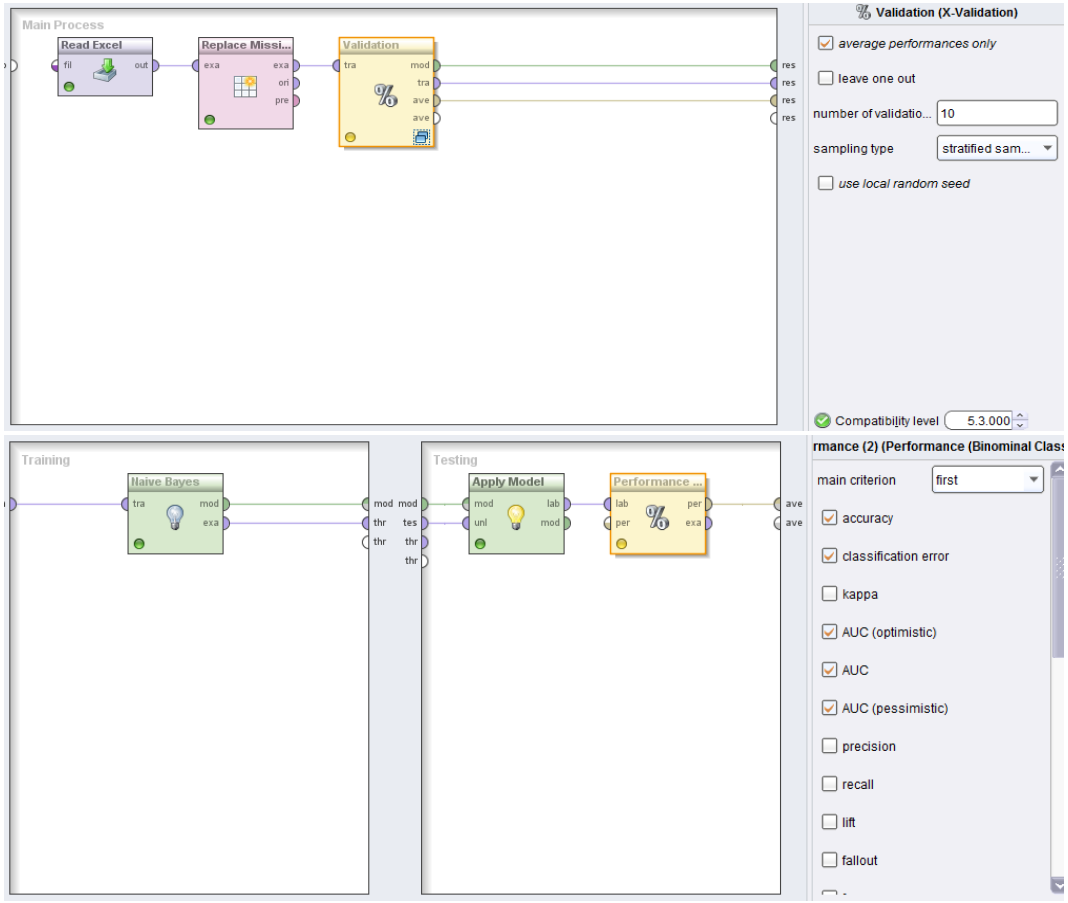
Sedangkan untuk mengisi *missing value* maka digunakan operator *Replace Missing Value* adalah *average* yang mempunyai arti data dengan *missing value* akan diisi dengan nilai rata-rata. Berikut ini merupakan desain pada Rapidminer 5.3 :



Gambar 4.2 *Replace missing value*

4.4.1 Algoritma *Naive Bayes Classifier*

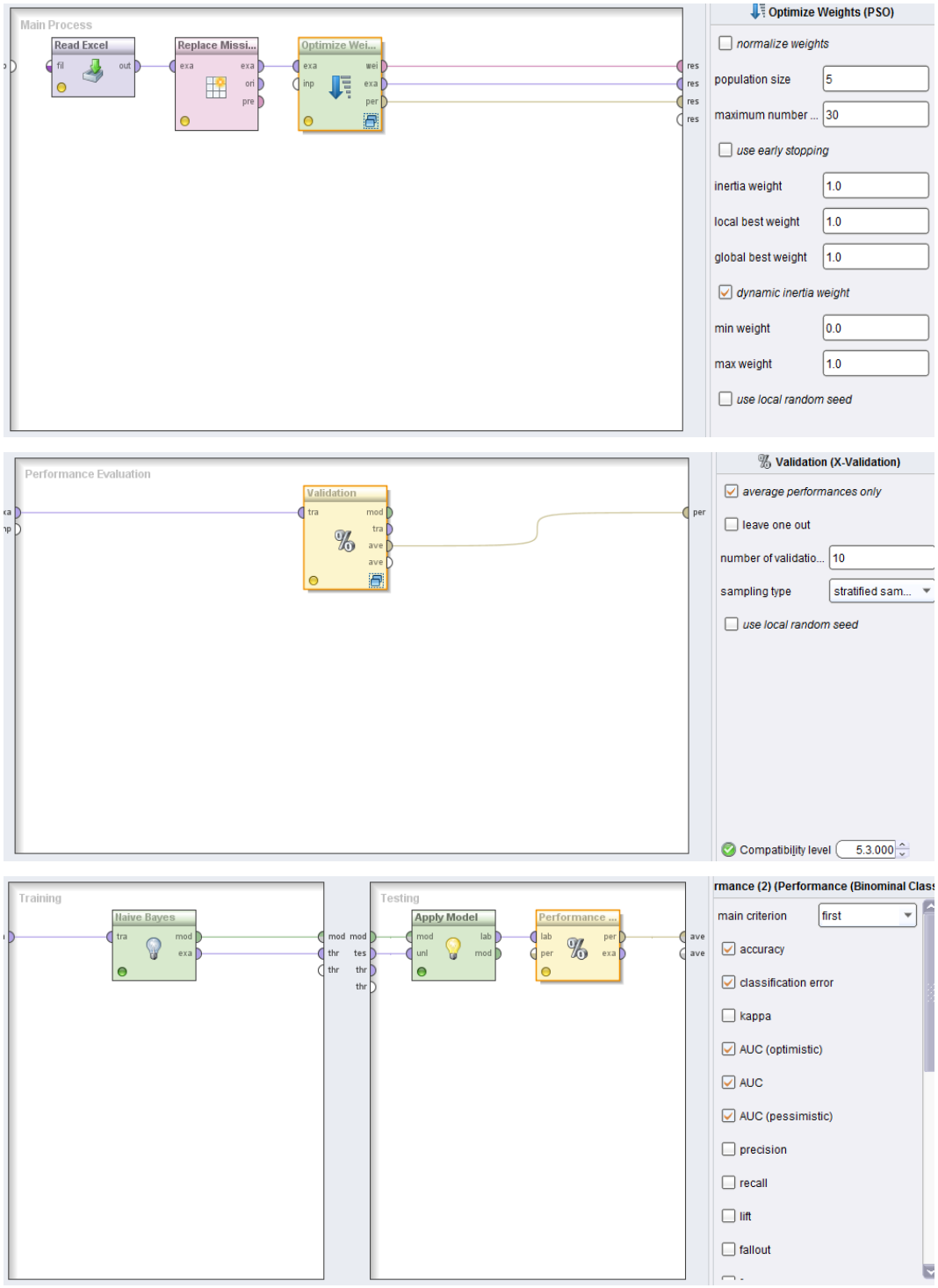
Eksperimen yang dilakukan dengan pendekatan data mining algoritma *Naive Bayes Classifier* menggunakan perangkat lunak Rapidminer 5.3 yaitu sebagai berikut :



Gambar 4.3 Model *Naive Bayes Classifier*

4.4.2 Algoritma *Particle Swarm Optimization-Naive Bayes Classifier*

Eksperimen yang dilakukan dengan pendekatan *Particle Swarm optimization* sebagai seleksi atribut pada algoritma *Naive Bayes Classifier* menggunakan perangkat lunak Rapidminer 5.3 yaitu sebagai berikut :

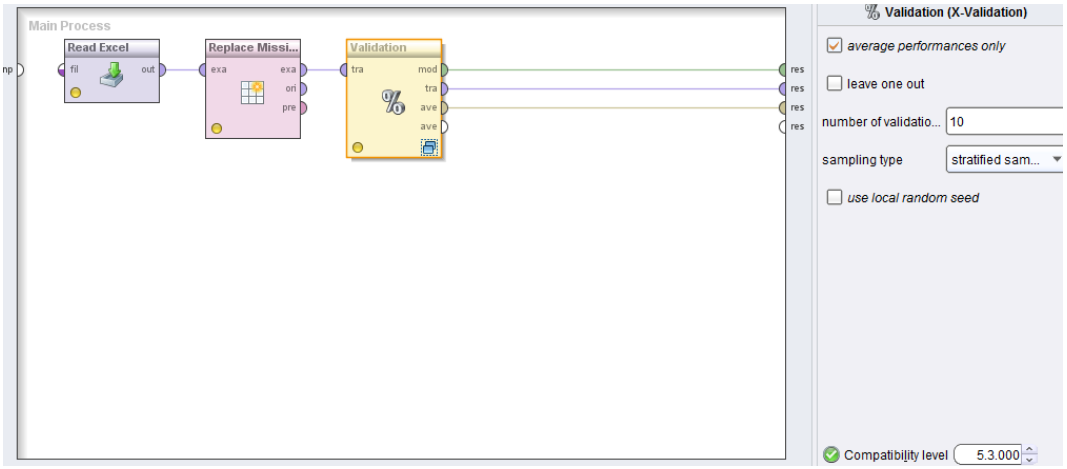


Gambar 4.4 Model *Particle Swarm Optimization - Naive Bayes Classifier*

4.5 Pengukuran

4.5.1 Algoritma *Naive Bayes Classifier*

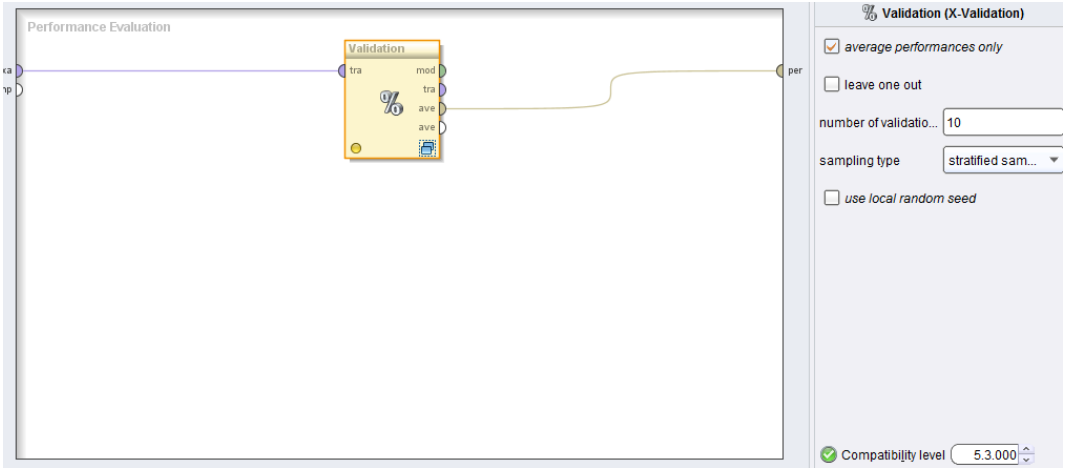
Berikut ini merupakan model untuk pengujian performansi algoritma *Naive Bayes Classifier* pada Rapidminer 5.3 :



Gambar 4.5 Validasi model *Naive Bayes Classifier*

4.5.2 Algoritma *Particle Swarm Optimization - Naive Bayes Classifier*

Berikut ini merupakan model untuk pengujian algoritma *Particle Swarm Optimization - Naive Bayes Classifier* pada Rapidminer 5.3 dengan 10-*fold cross validation.*



Gambar 4.6 Validasi model *Particle Swarm Optimization - Naive Bayes Classifier*

4.6 Hasil Analisa

4.6.1 Algoritma *Naive Bayes Classifier*

Hasil pendekatan model *Naive Bayes Clasifier* pada penelitian ini adalah sebagai berikut :

Tabel 4.4 Hasil performansi pendekatan *Naive Bayes Classifier*

Akurasi	Classification Error	AUC
85,08%	14,92%	0,841

4.6.2 Algoritma *Particle Swarm Optimization - Naive Bayes Classifier*

Hasil seleksi atribut dengan menggunakan *Naive Bayes Classifier* berbasis *Particle Swarm Optimization* adalah sebagai berikut :

Tabel 4.5 Hasil seleksi atribut *Particle Swarm Optimization*

ATRIBUT	WEIGHT
REGIONAL	0
WITEL	1
KWADRAN	0
PAKET EXISTING	0
R2BB	0
TYPE JARINGAN	0.9261941575329261
ZONA	0
MIGRASI	0

Tabel diatas merupakan hasil dari seleksi atribut menggunakan *Particle Swarm Optimization* dimana dari hasil menunjukkan bahwa atribut Regional, Kwadran, Paket Eksisting, R2BB, Zona, Migrasi bernilai 0. Sedangkan atribut Witel dan Tipe Jaringan bernilai 1 dan 0,926. Hal ini menunjukkan bahwa atribut-atribut yang bernilai lebih dari 0 terkait terhadap klasifikasi suatu data sehingga digunakan sebagai atribut prediksi klasifikasi pelanggan telepon dan speedy *cross*

seling produk Indihome dan sebaliknya atribut-atribut yang bernilai 0 tidak memiliki kaitan terhadap klasifikasi suatu data dan dapat dihilangkan untuk mempercepat kinerja model. Hal ini memiliki arti bahwa dataset yang semula terdapat 8 atribut prediktor menjadi 2 atribut prediktor.

Hasil pendekatan *Naive Bayes Clasifier* dengan seleksi atribut *Particle Swarm Optimization* pada model adalah sebagai berikut :

Tabel 4.6 Hasil performansi pendekatan *Particle Swarm Optimization-Naive Bayes Clasifier*

Akurasi	Classification Error	AUC
89,31%	10,69%	0,845

4.7 Pembahasan

4.7.1 Peningkatan akurasi

Berdasarkan analisa hasil evaluasi data mining untuk klasifikasi *cross selling* dengan algoritma *Naive Bayes* dan *Particle Swarm Optimization – Naive Bayes* maka dapat dirangkum sebagai berikut :

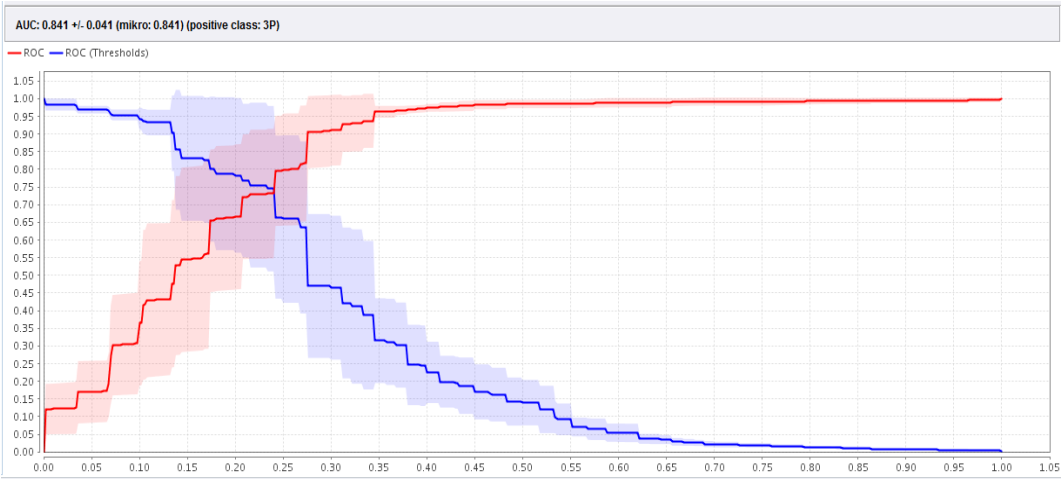
Tabel 4.7 Nilai perbandingan performansi

	NB	PSO - NB
Accuracy	85,08 %	89,31%
Classification Error	14,92 %	10,69%
AUC	0.841	0,845

Tabel diatas menunjukan bahwa dengan pendekatan *Naive Bayes Classifier* menghasilkan nilai akurasi sebesar 85,08% sedangkan dengan pendekatan *Particle Swarm Optimization – Naive Bayes Classifier* terdapat peningkatan nilai akurasi menjadi 89,31%.

Confusion Matrix

Nilai akurasi dengan pendekatan *Naive Bayes Classifier* adalah = 85,08% dapat dihitung dari hasil *Confusion Matrix* dimana *true positif* (tp) bernilai 534, *false negatif* (fn) bernilai 78, *false positif* (fp) bernilai 53 dan *true negatif* (tn) bernilai 213



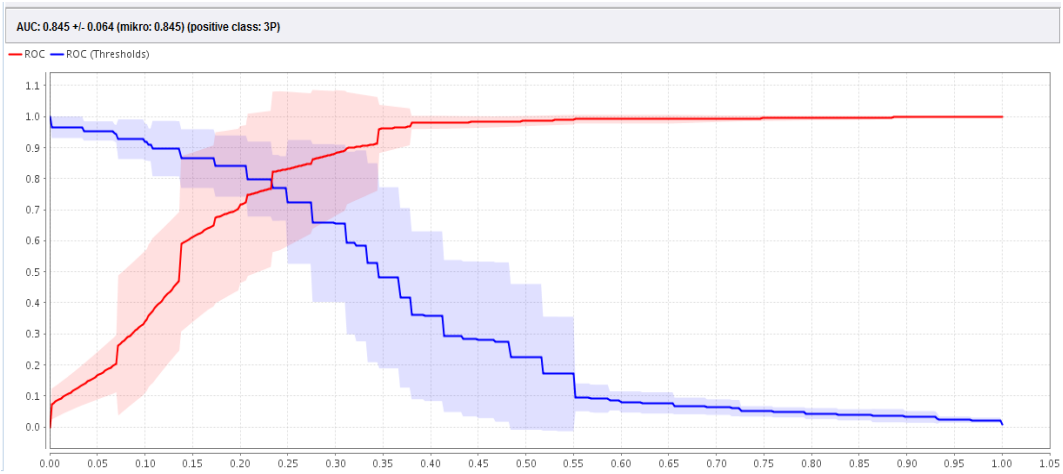
Gambar 4.7 *ROC Curve Naive Bayes Classifier*

ROC Curve

Hasil kurva ROC menujukkan arah garis melintang dari kiri bawah ke kanan atas didalam grafik dengan nilai AUC 0.841 sehingga dapat disimpulkan bahwa model *Naive Bayes* ini merupakan model klasifikasi yang Baik bahwa akurasi 0.80 – 0.90 = *Good classification.*

Sedangkan untuk pendekatan *Particle Swarm Optimization - Naive Bayes Classifier* menunjukan nilai akurasi dan AUC sebagai berikut

$$\text{Akurasi} = \frac{568 + 216}{568 + 216 + 75 + 19} = 0{,}8929$$



Gambar 4.8 *ROC Curve Particle Swarm Optimization - Naive Bayes Classifier*

ROC Curve

Sedangkan hasil kurva ROC menujukkan arah garis melintang dari kiri bawah ke kanan atas didalam grafik dengan nilai AUC 0.845 sehingga dapat disimpulkan bahwa model *Particle Swarm Optimization - Naive Bayes Classifier* ini merupakan model klasifikasi yang Baik bahwa akurasi 0.80 – 0.90 = *Good classification*.

4.7.2 Uji *T-test*

Prosedur *T-test* dilakukan untuk melihat perbedaan antara hasil algoritma *Naive Bayes Classifier* dan *Particle Swarm Optimization - Naive Bayes Classifier*. Berikut ini merupakan hasil uji statistik *T-test* :

T-Test Significance

	0.893 +/- 0.036	0.851 +/- 0.035
0.893 +/- 0.036		**0.017**
0.851 +/- 0.035		

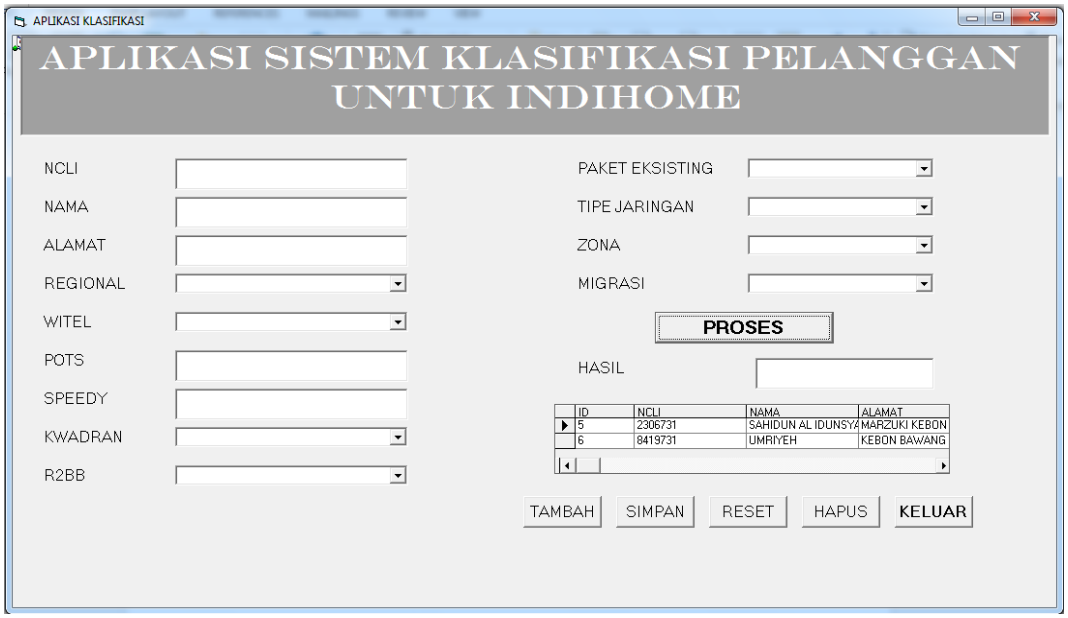
Probabilities for random values with the same result.
Bold values are smaller than alpha=0.050 which indicates a probably significant difference between the actual mean values!

Gambar 4.9 Uji T-test

Dari hasil pengujian dengan T-test diperoleh nilai $\alpha < 0.05$ yaitu 0,017 yang menunjukkan bahwa terdapat perubahan yang signifikan terhadap dua algoritma klasifikasi sehingga algoritma *Particle Swarm Optimization - Naive Bayes Classifier* merupakan algoritma yang signifikan untuk *data mining* klasifikasi ini.

4.8 Penerapan

Berdasarkan hasil penelitian yang telah dilakukan untuk data mining klasifikasi pelanggan telepon dan speedy eksisting PT Telkom Jakarta dalam *cross selling* produk Indihome maka diperoleh *rule* terbaik untuk diterapkan dalam pembuatan interface dengan algoritma *Naive Bayes Classifier* yang dapat memudahkan pemasaran PT. Telkom Jakarta. Interface yang digunakan dalam penelitian ini yaitu dengan menggunakan Microsoft Visual Basic 6.0.



Gambar 4.6 GUI Sistem Klasifikasi *Cross Selling* Produk Indihome

BAB 5 PENUTUP

5.1 Kesimpulan

Penelitian ini menggunakan *data mining* klasifikasi dengan algoritma *Naive Bayes* dan *Particle Swarm Optimization* sebagai seleksi atribut. Dari hasil penelitian diperoleh :

1. Akurasi yang dihasilkan algoritma *Naive Bayes Classifier* yaitu 85,08% , *classification error* = 14,92% dan AUC = 0,841
2. Akurasi yang dihasilkan algoritma *Naive Bayes Classifier* dengan *Particle Swarm Optimization* sebagai seleksi atribut meningkat menjadi yaitu 89,31% , *classification error* = 10,69% dan AUC = 0,845. Hal ini menunjukkan bahwa algoritma *Naive Bayes* dapat meningkatkan akurasinya dengan menggunakan *Particle Swarm Optimization* sebagai seleksi atribut.

Dengan demikian algoritma *Naive Bayes* dengan *Particle Swarm Optimization* sebagai seleksi atribut dapat memberikan pemecahan untuk permasalahan klasifikasi pelanggan *cross selling*.

5.2 Saran

Meskipun pada penelitian algoritma *data mining* ini menghasilkan *Naive Bayes* dengan *Particle Swarm Optimization* sebagai seleksi atribut dapat meningkatkan akurasi namun ada beberapa penambahan yang sebaiknya dilakukan, antara lain adalah:

a. Menambahkan lagi beberapa algoritma klasifikasi *data mining* untuk dikomparasikan seperti *Decision Tree, Logistic Regression, Artificial Neural Network* (ANN), *Support Vector Machines* (SVM) dan lain lain.
b. Memperbanyak jumlah atribut dan memanfaatkan metode penyeleksian atribut seperti *Chi-Square* dan lain-lain.

DAFTAR REFERENSI

Ahn, H. et al., 2011. Facilitating cross-selling in a mobile telecom market to develop customer classification model based on hybrid data mining techniques. *Expert Systems with Applications*, 38(5), pp.5005–5012.

Balaji, S. & K. Srivatsa, S., 2012. Naïve Bayes Classification Approach for Mining Life Insurance Databases for Effective Prediction of Customer Preferences over Life Insurance Products. *International Journal of Computer Applications*, 51(3), pp.22–26.

Bramer, M., 2007. *Principles of Data Mining, ser. Undergraduate Topics in Computer Science.*

Correa, E.S., Freitas, A. a. & Johnson, C.G., 2008. Particle Swarm for Attribute Selection in Bayesian Classification: An Application to Protein Function Prediction. *Journal of Artificial Evolution and Applications*, 2008, pp.1–12.

Elliott, A.C. & Woodward, W. a., 2007. Comparing One or Two Means Using the t -Test. *Statistical Analysis Quick Reference Guidebook: With SPSS examples*, pp.46–76.

Elsalamony, H. a, 2014. Bank Direct Marketing Analysis of Data Mining Techniques. , 85(7), pp.12–22.

Escalante, H.J., 2009. Particle Swarm Model Selection. , 10, pp.405–440.

Fayyad, U., Piatetsky-Shapiro, G. & Smyth, P., 1996. From data mining to knowledge discovery in databases. *AI magazine*, pp.37–54.

Ferdousy, E.Z., Islam, M.M. & Matin, M.A., 2013. Combination of Naïve Bayes Classifier and K-Nearest Neighbor (cNK) in the Classification Based Predictive Models. *Computer and Information Science*, 6(3).

Ghanad, N.K. & Ahmadi, S., 2015. Combination of PSO Algorithm and Naive Bayesian Classification for Parkinson Disease Diagnosis. , 4(4), pp.119–125.

Gorunescu, F., 2011. *Data mining: concepts and techniques*.

Harb, H.M., 2014. Feature Selection on Classification of Medical Datasets based on Particle Swarm Optimization. , 104(5), pp.14–17.

Jaroszewicz, S., 2008. Cross-selling models for telecommunication services. *Praca wys\lana do JTIT*, (December).

Jiawei, H. & Micheline, K., 2006. Data Mining Concepts and Techniques. *The Morgan Kaufman Series in Data Management systems*, p.770.

Kabir, M.F.M. et al., 2011. Enhanced Classification Accuracy on Naive Bayes Data Mining Models. *International Journal of Computer Applications (0975–8887) Volume*, 28(3), pp.9–16.

Karim, M. & Rahman, R.M., 2013. Decision Tree and Naïve Bayes Algorithm for Classification and Generation of Actionable Knowledge for Direct Marketing. *Journal of Software Engineering and Applications*, 06(04), pp.196–206.

Kennedy, J. & Eberhart, R., 1995. Particle swarm optimization. *Neural Networks, 1995. Proceedings., IEEE International Conference on*, 4, pp.1942–1948 vol.4.

Kothari, C., 2004. *Research methodology: methods and techniques.*

Larose, D.T., 2005. *Discovering Knowledge in Data an introduction to data mining.*

Li, J., Ding, L. & Li, B., 2014. A Novel Naive Bayes Classification Algorithm Based on Particle Swarm Optimization. , pp.747–753.

Lin, S.-W. et al., 2008. Particle swarm optimization for parameter determination and feature selection of support vector machines. *Expert Systems with Applications*, 35(4), pp.1817–1824.

Madyatmadja, E.D. & Aryuni, M., 2014. Comparative Study of Data Mining Model for. , 59(2), pp.269–274.

Moro, S. & Laureano, R.M.S., 2011. Using Data Mining for Bank Direct Marketing: An application of the CRISP-DM methodology. *European Simulation and Modelling Conference*, (Figure 1), pp.117–121.

Ren, S., Lian, Y. & Zou, X., 2014. Incremental Naïve Bayesian Learning Algorithm based on Classification Contribution Degree. *Journal of Computers*, 9(8), pp.1967–1974.

Sharma, A. & Panigrahi, P., 2011. A neural network based approach for predicting customer churn in cellular network services. *International …*, 27(11), pp.26–31.

Sheng, L.K. & Wah, T.Y., 2011. A comparative study of data mining techniques in predicting consumers' credit card risk in banks. *African Journal of Business Management*, 5(20), pp.8307–8312.

Taha, A.M., Mustapha, A. & Chen, S.-D., 2013. Naive Bayes-guided bat algorithm for feature selection. *TheScientificWorldJournal*, 2013, p.325973.

Taheri, S. & Mammadov, M., 2013. Learning the naive Bayes classifier with optimization models. *International Journal of Applied Mathematics and Computer Science*, 23(4), pp.787–795.

Ting, S.L., Ip, W.H. & Tsang, A.H.C., 2011. Is Naïve bayes a good classifier for document classification? *International Journal of Software Engineering and its Applications*, 5(3), pp.37–46.

Vihar, V. et al., 2009. Knowledge Discovery from Bio-medical Data Using a Hybrid PSO / Bayesian Classifier. , 2(1), pp.364–371.

Visa, S. et al., 2011. Confusion matrix-based feature selection. *CEUR Workshop Proceedings*, 710, pp.120–127.

Witten, I.H., Frank, E. & Hall, M. a., 2011. *Data Mining: Practical Machine Learning Tools and Techniques, Third Edition.*

Zaidi, N. & Cerquides, J., 2013. Alleviating Naive Bayes attribute independence assumption by attribute weighting. *The Journal of Machine …*, 14, pp.1947–1988.